Uses and Limitations of Multivariate Analyses in Analysis of Intercropping Experiments

by

W.T. Federer, Biometrics Unit, Cornell University, Ithaca, NY 14853

and

B.R. Murty, Expert, Int'l. Atomic Energy Agency, United Nations, Maracaibo, Venezuela

BU-858-M*

September 1984

Abstract

Since summarization of data from intercropping experiments involves a linear combination of crop yields, it would appear at first glance that multivariate analyses would be ideally suited for this situation. Linear combinations of values of crops, of total calories of crops, of total protein contents of crops, and of land utilization of crops are among some of the linear combinations which have considerable utility in summarizing data from intercropping experiments. From multivariate analyses, canonical variables based upon the criterion of maximum discriminating ability can be obtained when all crops are present in a mixture of crops. For v crops taken k at a time, such canonical variables are not obtainable using presently available theory. Even if they were, it is not certain that they would have any general utility for interpretational purposes. Areas of further research in multivariate analysis are discussed. These results are necessary in order to utilize multivariate analyses for intercropping investigations.

---

* In the technical report series of the Biometrics Unit, Cornell University, Ithaca, N.Y. 14853, U.S.A.

Uses and Limitations of Multivariate Analyses for Analysis of Intercropping Experiments

by

W.T. Federer, Biometrics Unit, Cornell University, Ithaca, NY 14853

and

B.R. Murty, Expert, Int'l. Atomic Energy Agency, United Nations, Maracaibo, Venezula

BU-858-M* September, 1984

1. Introduction

Intercropping investigations involve the growing of two or more crops simultaneously or sequentially on the same plot of ground. These mixtures of crops have been found to be beneficial in several respects when compared to yield responses from crops grown alone, i.e., sole crops. One problem is how to combine the yield responses from several crops, which may have considerably different means and variances. For a farmer who considers crop values in monetary terms, one could put a monetary value on each crop and obtain a total monetary value per hectare or per acre for each of the mixtures and for the sole crops. The form would be:

$$v_1C_1 + v_2C_2 + v_3C_3 \cdots + v_kC_k, \tag{1}$$

where v_i = value of ith crop, C_i = yield of ith crop in kilograms/hectare, say, and $i = 1, 2, \cdots, k$ = the number of crops in the mixture. Such a linear combination as (1) would provide a useful statistic for a grower of

---

* In the technical report series of the Biometrics Unit, Cornell University, Ithaca, N.Y. 14853, U.S.A.

crops. Likewise, if the farmer were interested in the total production of calories for his family needs, again we could use a linear combination of the form:

$$c_1C_1 + c_2C_2 + c_3C_3 + \cdots + c_kC_k \tag{2}$$

where c_i is the calorie conversion coefficient for the ith crop. A similar linear combination could be used for protein conversion.

If the farmer were interested in land use efficiency, he could use a linear combination of yield responses for each crop of the form:

$$C_1/C_{1s} + C_2/C_{2s} + C_3/C_{3s} + \cdots + C_k/C_{ks} \tag{3}$$

where C_{is} is the yield of the ith crop when grown as a sole crop. The coefficients v_i and c_i would normally be taken as constants with the C_i being the only random variables. Then if $C_1, \cdots, C_k$ have a k-variate multivariate normal distribution (1) and (2) have a univariate normal distribution and no problems of statistical analyses are encountered over those found for sole cropping. If the C_{is} are also random variables, (3) would give a distribution which is a sum of Cauchy variables. However, if the C_{is} are constants, which might be the average yields of crops in that region for the past y years, then (3) also has a univariate normal distribution.

Now crop values and yields fluctuate from year to year but the ratios of prices, v_i/v_1 and of yields, C_{1s}/C_{is}, where v_1 is the value for a base crop and C_{1s} is the yield for a base crop, are relatively stable. Hence, one could use $\Sigma v_iC_i/v_1$ in place of (1) and $C_{1s}\ \Sigma\ C_i/C_{is}$ in place of (3). One could do likewise for (2). Let us call these "relative crop values," "relative calories," and "relative land use." In this form then, all linear combinations have a similar form,

$$C_1 + b_2C_2 + b_3C_3 + \cdots + b_kC_k \quad . \tag{4}$$

The above is also the form for a canonical variable obtained from a multivariate analysis. As we shall see this form is useful when comparing the various forms with each other and especially with the canonical variable obtained from a multivariate analysis.

We first consider a multivariate analysis for p characteristics on crop i. There are no difficulties encountered here which are due to inter-cropping. In the third section the case of c_1 lines of crop one and c_2 lines of crop two in all possible combinations is discussed. The yields of crop one are considered to be the first variable X_1 and the yields of the second crop are considered to be the second variable X_2. A series of interpretational problems are encountered for this situation. The present state of multivariate theory does not allow for solution of these problems and the basic concept of multivariate analysis of variance (MANOVA), which is to obtain a canonical variable which has maximum discriminating ability, is questioned. In Section 4, the case of p crops in mixtures of k crops, k = 1, 2, $\cdots$, p, is considered. Even if present multivariate theory were applicable in Section 3, it is not sufficient to handle the problems encountered in Section 4 even if k is not allowed to vary. Section 5 considers the situation for which there are c_i lines for the ith crop and these are mixtures of k of p crops. Even if each of the line combinations from the $\Pi_{i=1}^{p} c_i = v =$ total number of line combinations are considered separately, we are left with the difficulties encountered in Section 3. Finally, some areas of research to extend present multivariate theory and analyses are discussed in the last section.

2. Multivariate Analyses for p Characteristics of One Crop

When there are p characteristics or variables for a single crop, and provided that the necessary normality assumptions are satisfied, standard multivariate procedures may be utilized. The purpose here would be to obtain linear combinations, canonical variables, of the p variables which discriminate among the treatments in the experiment. It would be hoped that one or two canonical variables would summarize the whole of the information for the p variables. Also, the relative importance of the p variables in discriminating among the treatments could be assessed.

Thus, when using multivariate procedures for the purpose described here, there appear to be no problems in summarizing information from data derived from intercropping investigations. For example, suppose that an intercropping experiment on maize-beans mixtures in combination and with maize and beans grown alone, sole crops, was conducted. Suppose that there were two lines of maize, X and Y, and four lines of beans, A, B, C, and D, then the 14 treatments (marked X) in the experiment would be:

	beans				
Maize	A	B	C	D	Sole
X	x	x	x	x	x
Y	x	x	x	x	x
Sole	x	x	x	x	

There are eight mixtures plus two maize sole crops plus four bean sole crops. If M_1 = total weight of maize grain for an experimental unit and M_2 = number of ears per maize plant, there would be two characteristics for maize and a bivariate analysis could be performed for the ten treatments involving maize, i.e., the eight mixtures and the two sole crops for maize. Likewise, for the four variables measured on beans, B_1 = number of beans per pod, B_2 = number of pods per plant, B_3 = grain weight of 100 beans, and B_4 = grain weight per experimental unit, one could perform a multivariate analysis using the 12 treatments involving beans. All four variables or some subset of the four variables could be included in the multivariate analysis.

3. Multivariate Analysis for Mixtures of c_1 Lines of Crop One With c_2 Lines of Crop Two

As was stated in the Introduction, a goal of statistical analyses for intercropping investigations is to obtain some linear combination of the yields from both crops in a mixture. If we let crop one be variable one, say X_1, and crop two yields be variable two, say X_2, a bivariate analysis of variance may be performed (see Pearce and Gilliver, 1978 and 1979; Mead and Riley, 1981). As mentioned by the above authors, one serious problem here is heterogeneity of covariance matrices. The variance and covariance of a line of one crop may vary from line to line of the second crop. The lines of any one crop may, and often do, have heterogeneous covariance structures. To overcome this, it is suggested that single degree of freedom contrasts be made from the treatment sum of squares. It may also be necessary to compute an individual error variance for each single degree of freedom contrast. This can easily be accomplished for completely randomized and randomized complete blocks designs.

One problem that appears to be unsolvable using presently available multivariate theory is how to compare sole crop yields for one crop, say M_s, with a linear combination of two crops in a mixture, say $M_m + bB_m$. If b is a ratio of crop values, e.g., this can easily be done. If b is a ratio of yields of sole crop yields in farmers fields, say $b = M_s/B_s$, then again, one can compare sole crop and mixture yields. However, if b is a coefficient obtained from a bivariate analysis to form a canonical variable, multivariate theory is not sufficiently advanced to allow comparisons between M_s and $M_m + bB_m$. This is a serious deficiency in the theory as it is usually desired to compare sole cropping with mixture yields in intercropping investigations.

A second problem arising from the use of multivariate analyses for data from intercropping investigations is the usefulness and interpretation of a canonical variable of the form $M_m + bB_m$. It appears that this canonical variable obtained from a bivariate analysis is useless for interpretation of intercropping data on mixtures of two crops. The criterion used to obtain the canonical variable, i.e., the linear combination producing maximum discriminating ability among treatments has no meaning. To illustrate consider a set of data obtained from an experiment designed as a randomized complete block with four blocks and including the eight maize and bean mixtures described above. A bivariate analysis of variance is given in Table 1. For the discussion, ignore the fact that data were missing for beans in one of the bean-maize mixtures. Carrying through the computations, it was found that b in $M_m + bB_m$ was 38. Now a meaningful b in terms of prices of maize and beans is between three and seven. A meaningful b in terms of land use, is in the same range. These values are not even close to the one obtained in the bivariate analysis. To put this into perspective, Figure 1 has been prepared where the canonical correlations are computed for various values of b = R. The canonical correlation is computed from the formula of treatment sum of squares for the canonical variable divided by the treatment plus error sums of squares. The formula is

$$S = \frac{M_{tt} + R\,C_{tt} + R^2 B_{tt}}{M_{tt} + M_{ee} + R(C_{tt} + C_{ee}) + R^2(B_{tt} + B_{ee})}, \qquad (4)$$

where M_{tt} and M_{ee} are the treatment and error sums of squares, respectively, for maize yields in the mixtures and similarly for B_{tt} and B_{ee} for beans, and C_{tt} and C_{ee} are the cross products of maize and bean yields for

Table 1. Bivariate analysis of variance for mixtures of maize and beans with crop yields as variables.

Source of variation	Degrees of freedom	Sums of products	Covariance matrix
Total	32[1]	$\begin{bmatrix} 60,936.45 & 628,076 \\ - & 7,097,198 \end{bmatrix}$	-
Correction for mean	1	$\begin{bmatrix} 57,231.903 & 625,055.44 \\ - & 6,826,512.5 \end{bmatrix}$	-
Blocks	3	$\begin{bmatrix} 702.003 & 1,966.65 \\ - & 45,284.75 \end{bmatrix}$	$\begin{bmatrix} 234.00 & 655.55 \\ - & 15,094.92 \end{bmatrix}$
Treatments	7	$\begin{bmatrix} 1,540.980 & 1,810.06 \\ - & 66,340.00 \end{bmatrix}$	$\begin{bmatrix} 220.14 & 258.58 \\ - & 9,777.14 \end{bmatrix}$
Remainder	21[1]	$\begin{bmatrix} 1,461.564 & -756.15 \\ - & 159,060.75 \end{bmatrix}$	$\begin{bmatrix} 73.078 & -37.808 \\ - & 7,574.32 \end{bmatrix}$

[1] less one for first variable bean yields

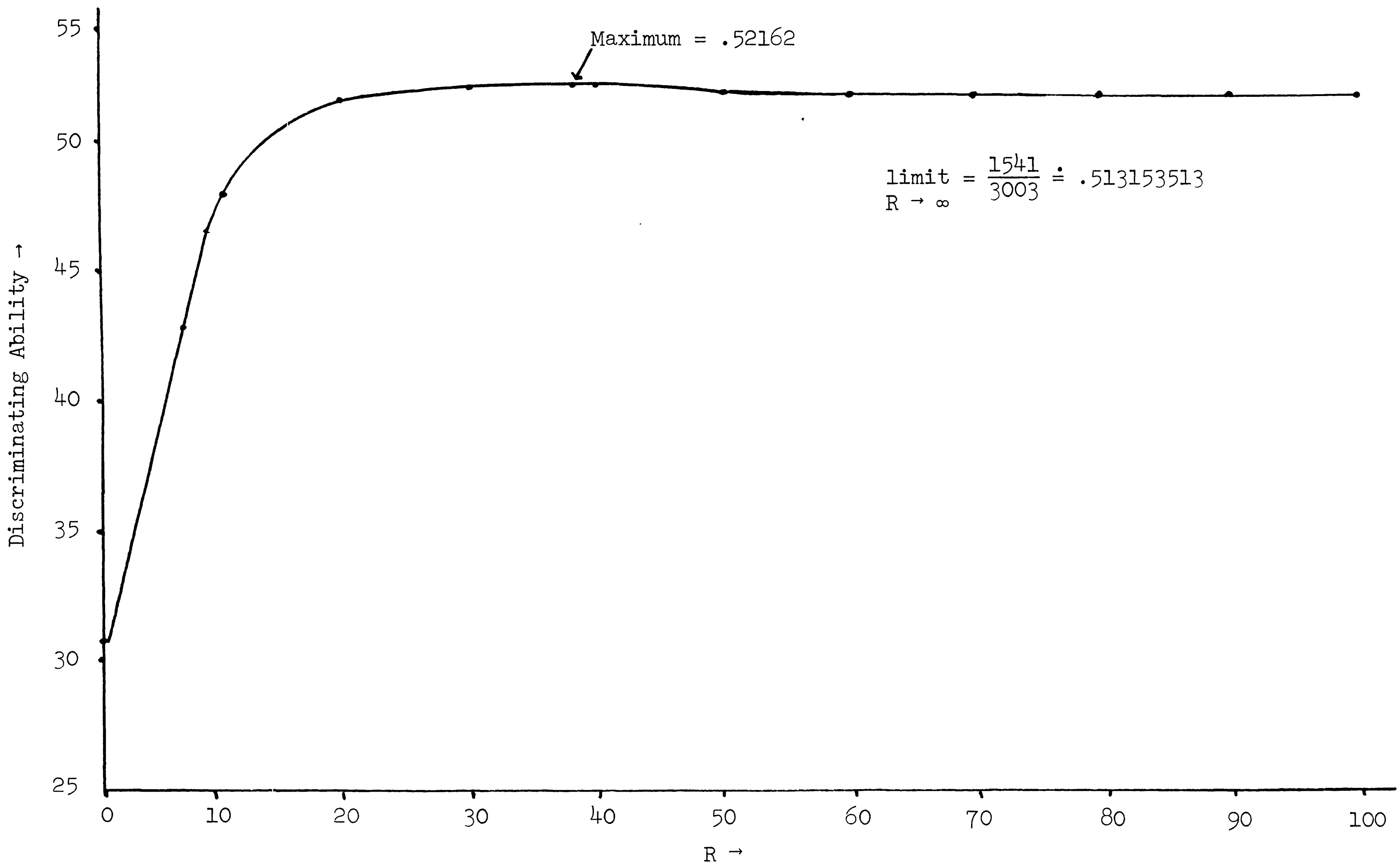

Figure 1. Treatment/(treatment + error) sums of squares for $0 \leq R \leq 100$.

treatments and for error, respectively. The canonical correlation measures discriminating ability of the canonical variable. S has been computed for values of $0 \leq R \leq 100$. The range of $3 \leq R \leq 7$, or even $2 \leq R \leq 12$, has a practical interpretation whereas values of R in the range of 35 to 40 have no meaning, throwing considerable doubt on the criterion of "maximum discriminating ability." One other point of interest is the relative flatness of the curve for S when $R > 20$, with a limiting value close in value to the maximum at 38. Because of this problem, it would appear that this type of multivariate analysis is inappropriate for analyzing data from intercropping experiments. One should also note that "maximum discriminating ability" is not invariant with respect to differences among maize or among bean lines.

4. Multivariate Analyses for p Crops in Mixtures of k Crops With One Line Per Crop

In the previous section, two crops were considered. There were c_1 lines of crop one and c_2 lines of crop two. Here we consider the case where there are mixtures of k crops from among p crops, $k < p$. Also, here we consider that there is only one line per crop. In the next section the case of c_i lines on the ith crop is discussed. To illustrate consider that four crops are to be included in an investigation wherein mixtures of three of the four crops are studied. All possible mixtures of three crops, for A being crop one, B crop two, C crop three, and D crop four, would be:

Mixture			
1	2	3	4
A	A	A	B
B	B	C	C
C	D	D	D

There also could be four sole crops, six mixtures of two crops, i.e., AB, AC, AD, BC, BD, and CD, and one mixture of four crops, ABCD. For the above discussed set of mixtures of three crops, let us designate the yield response on crop i as X_i, i = A, B, C, D. The measurements on the four variables (crops) in one of the r complete blocks for mixtures of three would be:

	Variable (crop)			
mixture	X_A	X_B	X_C	X_D
1	x	x	x	
2	x	x		x
3	x		x	x
4		x	x	x

where x denotes that an observation for the variable is available and a blank means that no value for the variable was obtained. The design for the responses in this case would be a balanced incomplete block design with design parameters $v^* = 4$, $b^* = 4$, $r^* = 3$, $k^* = 4$, and $\lambda^* = 2$. In computing a sum of squares for a variable X_i, rr^* measurements would be used. To compute a sum of cross products $r\lambda^*$ observations would be used for any pair of crops. Let T_{ii} and E_{ii} represent a treatment and error sum of squares for variable X_i, respectively, and let T_{ij} and E_{ij}, $i \neq j = 1, 2, \cdots p$, represent the cross products between crops i and j. Testing in a multivariate analysis involves determinants of matrices composed of sums of squares and cross products, of the form:

$$\frac{\begin{vmatrix} E_{11} & E_{12} & E_{13} & \cdots \\ E_{12} & E_{22} & E_{23} & \\ E_{13} & E_{23} & E_{33} & \end{vmatrix}}{\begin{vmatrix} E_{11}+T_{11} & E_{12}+T_{12} & E_{13}+T_{13} & \cdots \\ E_{12}+T_{12} & E_{22}+T_{22} & E_{23}+T_{23} & \\ E_{13}+T_{13} & E_{23}+T_{23} & E_{33}+T_{33} & \\ \vdots & & & \end{vmatrix}} . \tag{5}$$

Since the E_{ii} are computed from rr^* reponses and the E_{ij} from $r\lambda^*$ responses, one could use a conservative approach and multiply each E_{ii} by λ^*/r^*. This would put sums of squares and sums of cross products on the same basis in order that a test could be used. Such a procedure would not utilize all the information in the data but would allow use of standard multivariate analyses. It should be noted that rr^* and $r\lambda^*$ can be quite different in value. For example, let $v^* = 25$, $k^* = 5$, $r^* = 6$, $b^* = 30$, and $\lambda^* = 1$. For this case, $\lambda^*/r^* = 1/6$, or $rr^* = 6r\lambda^*$. Even for $v^* = 7 = b^*$, $k^* = 3 = r^*$, and $\lambda^* = 1$, $rr^* = 3r\lambda^*$. For $v^* = 27$, $r^* = 13$, $k^* = 3$, $b^* = 117$, and $\lambda^* = 1$, $rr^* = 13\ r\lambda^*$. The number of crops and the size of the mixture k^* of crops determines the disparity in rr^* and $r\lambda^*$ values.

At first glance, it might appear that the paper by Srivastava (1968) would be useful in analyzing incomplete responses such as described above. A study of the paper shows this not to be the case, and it is not known how to use the results for situations encountered in intercropping investigations. Srivastava (1968) considers the following situation. For t treatments in a block design on which p responses are to be studied, some or all of the responses are obtained for a subset of the blocks. To illustrate let $t = 4$, $r = 4$, $k = 3$, $b = 4$, and $\lambda = 2$ be the design on which a set of responses is measured. Let the design on the responses also be a blocked design for the $p = 3$ response variables. For this design let $S_1 = (X_1, X_2)$, $S_2 = (X_1, X_3)$, $S_3 = (X_2, X_3)$, and $S_4 = (X_1, X_2, X_3)$, i.e., two of the three responses are obtained in three sets of b blocks and all three responses are obtained in one set of b blocks. This would require a total of $4(4) = 16$ blocks of size $k = 3$ experimental units. If each S_h, $h = 1, 2, 3, 4$, is included twice as in Srivastava's (1968) example, the plan in Table 2 would result. There would be a total of 32 blocks of size $k = 3$

Table 2. Layout for Srivastava (1968) example

Set	Variable	Block 1			Block 2			Block 3			Block 4		
		treatment 1	treatment 2	treatment 3	treatment 1	treatment 2	treatment 4	treatment 1	treatment 3	treatment 4	treatment 2	treatment 3	treatment 4
S_1	V_1												
	V_2												
S_2	V_1												
	V_2												
S_3	V_1												
	V_3												
S_4	V_1												
	V_3												
S_5	V_2												
	V_3												
S_6	V_2												
	V_3												
S_7	V_1												
	V_2												
	V_3												
S_8	V_1												
	V_2												
	V_3												

experimental units each for a total of 96 experimental units. The total number of observations would be $6(2)(12) + 2(3)(12) = 216$.

Purportedly the reason for not observing responses on all three variables on every one of the 96 experimental units was cost or lack of time. It would appear that a more reasonable and desirable procedure would be to use one third fewer blocks and measure all variables on each experimental unit. Then one could use standard multivariate procedures and not run into the difficulties discussed in the Srivastava (1968) paper.

Srivastava (1968) was interested in the treatments in the design. In intercropping the sets S_h are the treatments and are the items of interest in the investigation. Thus interest was on the column totals of Table 2, whereas the row totals of Table 2 represent the treatments in an intercropping investigation. A further study of the procedures described threw no light on how to adapt the procedures of the paper for the situation encountered.

5. Multivariate Analyses for p Crops in Mixtures of k Crops with c_i lines on Crop i

Consider the situation wherein there are p crops which will be grown in mixtures of k crops and that there are c_i lines on crop i, $i = 1, 2, \cdots, p$, $k < p$. For the $\binom{p}{k}$ combinations or some subset thereof, one could conduct k-variate analysis in the manner described in Section 3 for one pair of crops. The extention would be straightforward. The difficulties discussed for the bivariate case would carry over to the k-variate case.

If one wished to use a combined analysis over all $\binom{p}{k}$ combinations then all the difficulties described in Section 4 arise when one attempts a multivariate analysis. If the values v_i in (1), c_i in (2), or C_{si} in (3) are given or if the relative values b_i in (4) are known, then there is no difficulty in forming such a canonical variable or a relative canonical variable. The only difficulty that would arise would be in obtaining a range of these values. For two crops, it is easy to vary b in $X_1 + bX_2$. For three crops it is again not too difficult to vary b and c in X_1, $X_1 + bX_2$, $X_1 + cX_3$, $bX_2 + cX_3$, bX_2, cX_3, and $X_1 + bX_2 + cX_3$ where there would be sole crops, mixtures of two crops, and a mixture of all three crops to compare. One could look at the various levels of b at various levels of c. There would be $\ell_2\ell_3$ levels to compare the various treatments (mixtures) in the experiment. With p crops there would be $\ell_2\ell_3\cdots\ell_p$ total levels at which to compare the treatments in the experiment, where ℓ_i is the relative level (crop value, calories, and/or land use) for crop i. This many computations makes presentation and interpretation difficult. Hence, ℓ_i should be as small as possible. For Figure 1, we used $0 \le \ell_2 \le 100$, but such a range of values would be extremely tedious if one used $0 \le \ell_2 \le 100$, $0 \le \ell_3 \le 100$, $0 \le \ell_4 \le 100$, etc. Therefore, it is recommended that at most three values for each ℓ_i be used, i.e., a low value, an average value, and a high value. For calorie or protein conversions, it may not be necessary to vary the ℓ_i. For land use and for monetary value, these values should be varied.

6. Some Unresolved Problems in Multivariate Analyses

One problem raised in the preceding discussion was the comparison of sole crops with mixtures of two, three, $\cdots$, the comparison of mixtures of two with mixtures of three, four, etc. One method of doing this might be the following. Let α_i be the proportion of the variate in a multivariate distribution with $\Sigma_{i=1}^{p}\alpha_i = 1$. When the $\alpha_i = \alpha$, we have present multivariate theory. If one were able to generalize present multivariate theory, this would appear to be one way of making the above comparisons, though it must be realized that they may not be useful in a practical interpretation of data from intercropping experiments. It would solve the problem of comparing canonical variables with different numbers of variables.

The use of weight α_i for variable X_i in a multivariate normal distribution could be useful in analyzing replacement series data such as depicted in Figure 2. Here the proportion of one crop goes from zero to one while a second crop goes from one to zero. If the effect of a mixture was simply substitution of crops with no beneficial or no detrimental effect of the mixture, then the yields of the mixture would follow the solid line in Figure 2. If one developed the generalized multivariate normal as described above, it would be for this situation.

On the other hand, suppose that the mixture yields followed the dashed line above the solid line. This would indicate a beneficial effect of mixing two crops which is a usual result in intercropping if the crops are judiciously selected (Okigbo, 1981). If the curve were a smooth function, perhaps only one additional parameter in the generalized multivariate normal distribution would be required to handle the situation. However, a development of the theory for this situation would be useful.

A second line of development required in multivariate analysis is the one discussed in Sections 4 and 5. This has to do with the computation of sums of squares and cross products from unequal numbers of observations. A simple problem in this area is the distribution of the variance of the linear regression coefficient computed as β = covariance (x,y)/variance (x), where the covariance (x,y) has N_1 degrees of freedom and variance (x) has N_2 degrees of freedom. For the multivariate situation let us consider a specific case for $p = 3$, for sole crop yields X_{1s}, X_{2s}, and X_{3s}, for a

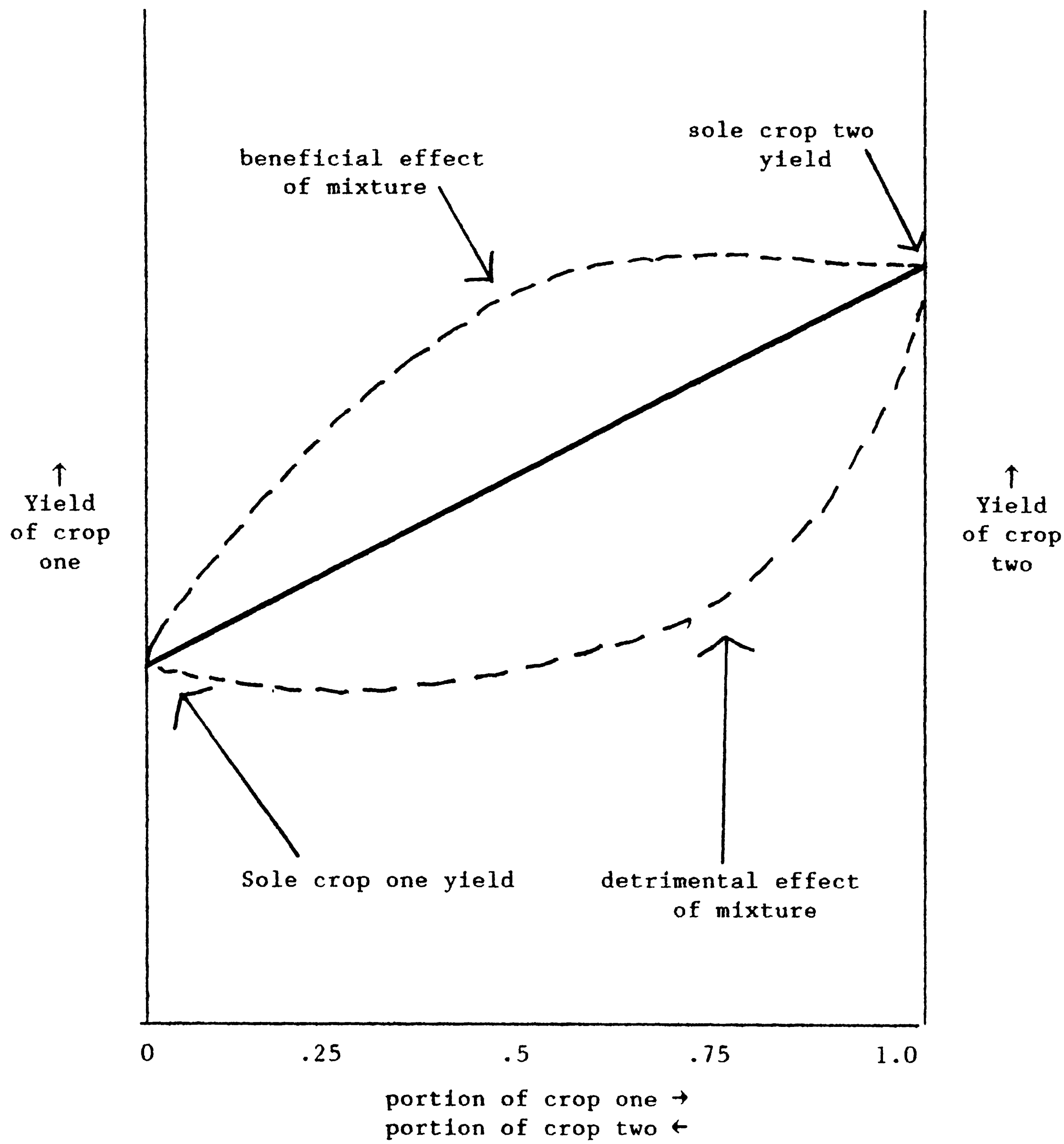

Figure 2. Replacement series for two crops

mixture of two crop yields X_1 and X_2, X_1 and X_3, X_2 and X_3 and for a mixture of all three crops X_1, X_2, and X_3. In a completely randomized design one may compute the following matrix of sums of squares and cross products:

$$\left[\begin{array}{ccc|ccc|ccc}
S_{11}(\text{sole}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & S_{22}(\text{sole}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & S_{33}(\text{sole}) & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & S_{11}(2) & S_{12}(2) & S_{13}(2) & 0 & 0 & 0 \\
0 & 0 & 0 & S_{12}(2) & S_{22}(2) & S_{23}(2) & 0 & 0 & 0 \\
0 & 0 & 0 & S_{13}(2) & S_{23}(2) & S_{33}(2) & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & S_{11}(3) & S_{12}(3) & S_{13}(3) \\
0 & 0 & 0 & 0 & 0 & 0 & S_{12}(3) & S_{22}(3) & S_{23}(3) \\
0 & 0 & 0 & 0 & 0 & 0 & S_{13}(3) & S_{23}(3) & S_{33}(3)
\end{array}\right]$$

Where $S_{ii}(2)$ is a sum of squares computed from the yields of crop i when it is in mixtures of two and computed from 2r observations, $S_{ii'}(2)$, $i \neq i'$, is a sum of cross products for crops i and i' computed from r observations, $S_{ii}(3)$ is a sum of squares for crop i computed from mixtures of three and obtained from r observations, $S_{ii}'(3)$, $i \neq i'$, is a sum of cross products computed from the r observations, from mixtures of three, and $S_{ii}(\text{sole})$ is a sum of squares computed from r sole crop yields in the r replicates.

There are only three variates but there is a nine by nine matrix with the population variances for the nine diagonals being different. Also, there are six different parameters for the covariances. The mean for crop X_i changes from sole to mixtures of two to mixtures of three. For this situation, comparisons of sole crop yields with mixtures of two and of three crops and comparisons of mixtures of two crops with a mixture of three crops are desired. One has only a three variate problem which turns into a situation resembling a nine variate problem.

In the above, if we considered only mixtures of two, the sums of squares would be computed from 2r observations and the sums of cross products from r observations. Now what is the distribution of error over treatment plus error sums of squares and cross products? One would be considering a quotient of the form given by equation (5). In general if there are $\binom{v^*}{k^*}$ combinations, a balanced incomplete block design with parameters v^*, k^*, $b^* = v^*!/k^*!(v^* - k^*)!$, $r^* = (v^* -1)!/(k^* -1)!(v^* -k^*)!$, and $\lambda^* = (v^* - 2)!/(k^* - 2)!(v^* - k^*)!$. The sums of squares from crops in mixtures of k would be computed from rr^* observations and the sums of cross products are computed from $r\lambda^*$ observations. How does the estimation and testing proceed for this situation?

A third problem that arises is the interpretation of results from an intercropping experiment when equation (1), (2), (3), or (4) is used as the first canonical variable. Then, one computes a multivariate analysis and obtains canonical variables after (4), say, the computational procedure needs to be detailed as well as determining whether or not this procedure produces any useful results (see, e.g., Burnaby, 1966).

In connection with the preceding, the criterion of maximum discriminating ability arises. From the example given it would appear that a canonical variable arising from the criterion of maximum discriminating ability has no practical interpretation in intercropping investigations. The question arises as to what other criteria should be used in analyzing data from intercropping experiments. Should we continue to cling solely to the criterion set forth by Fisher (1936, 1938, 1940) and Smith (1936) or should multivariate theory proceed using other criteria?

Another question that arises is the following. Suppose one has k canonical variables to summarize the information from p variables, k < p. Now should one do a second-stage multivariate analysis and treat the k canonical variables as a k-variate problem? Can one now obtain a linear combination of linear combinations and have a stage 2 canonical variable? Should one proceed to an s stage multivariate analysis? This problem arises in intercropping. Should one treat equation (1) as variate one, equation (2) as variate two, and equation (3) as variate three and conduct a three-variate multivariate analysis?

7. Literature Cited

Burnaby, T. (1966). Growth invariant discriminant functions and generalized functions. *Biometrics* 22, 96-110.

Fisher, R.A (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics 7,* 179-188.

Fisher, R.A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics 8,* 376-386.

Fisher, R.A. (1940). The precision of discriminant functions. *Annals of Eugenics 10,* 422-429.

Mead, R. and J. Riley (1981). A review of statistical ideas relevant to intercropping research (with discussion). *J. Royal Statistical Soc. (A) 144,* 462-509.

Okigbo, B.N. (1981). Evaluation of plant interactions and productivity in complex mixtures as a basis for improved cropping systems design. *ICRISAT, Proc., International Workshop on Intercropping,* pp 155-179.

Pearce, S.C. and B. Gilliver (1978). The statistical analysis of data from intercropping experiments. *J. Agric. Sci., Cambridge, 91,* 625-632.

Pearce, S.C. and B. Gilliver (1979). Graphical assessment of intercropping methods, *J. Agric. Sci.,* Cambridge 93, 51-58.

Smith, H.F. (1936-7). A discriminant function for plant selection. *Annals of Eugenics 7,* 240-250.

Srivastava, J.N. (1968). On a general class of designs for multiresponse experiments. *Annals Math. Stat. 39,* 1825-1843.